Expected Likelihood Support for Blind SIMO Channel Identification

Yuri I. Abramovich
Institute for Telecommunications Research
University of South Australia
Mawson Lakes, SA, 5095 Australia
Email: yuri.i.abramovich@gmail.com

Ben A. Johnson
Elec. Eng. & Comp. Sci. Dept.
Colorado School of Mines
Golden, CO 80401 USA
Email: ben.a.johnson@ieee.org

***Abstract*—SIMO channel identification problems arise in many practical applications, such as geolocation of HF sources propagated via the multi-layer ionosphere. In this case, memory of the channel (often modeled as a finite impulsive response (FIR) channel) makes the traditional assumptions on the channel estimation training samples as independent and identically distributed (i.i.d) invalid. This potentially precludes the use of statistical characteristics typically derived under the i.i.d. assumption, including the Expected Likelihood quality assessment technique. In this paper, we introduce a likelihood-like criteria for this circumstance and demonstrate the practical invariance properties of its distribution for the Expected Likelihood condition, met when the estimated parameters are statistically equivalent to the true ones.**

I. Introduction

In problems with multi-mode propagation of signals to an M-element antenna array, the M-variant "snapshot" vector $\mathbf{x}(t)$ observed at the moment t may be presented as:

$$\mathbf{x}(t) = \sum_{\tau\in\mathbb{Z}} \mathbf{h}(\tau)\mathbf{s}(t-\tau) \tag{1}$$

where $\{\mathbf{h}(\tau)\}_{\tau\in\mathbb{Z}}$ is the vector-valued impulse response, $\mathbf{h}(\tau) = [h_1(\tau), \ldots, h_M(\tau)]^T$, $\mathbf{s}(t)$ is the transmitted source waveform and $\mathbb{Z}$ is the domain of the channel impulse response.

Let us assume that the transfer function $\mathbf{h}(z)$

$$\mathbf{h}(\mathbf{z}) = \sum_{\tau\in\mathbb{Z}} \mathbf{h}(\tau)e^{-\mathbf{z}\tau} \tag{2}$$

corresponds to the causal FIR system of degree L (*i.e.* $\mathbf{h}(L) \neq 0$). Let the noisy output $\mathbf{y}$ of that system with snapshot $\mathbf{x}$ be

$$\mathbf{y}(t) = \mathbf{x}(t) + \mathbf{w}(t) \tag{3}$$

with spatially uncorrelated and temporally white additive noise $\{\mathbf{w}(t)\}_{\mathbf{t}\in\mathbb{Z}}$ of variance σ^2 independent of $\{\mathbf{x}(t)\}_{t\in\mathbb{Z}}$.

Then for a sequence of $(N+1)$ time observations at the output of the M-element antenna array, with

$$\begin{aligned} \mathbf{x}_N(t) &= [\mathbf{x}^T(t), \mathbf{x}^T(t-1), \ldots, \mathbf{x}^T(t-N)]^T \\ \mathbf{y}_N(t) &= [\mathbf{y}^T(t), \mathbf{y}^T(t-1), \ldots, \mathbf{y}^T(t-N)]^T \\ \mathbf{w}_N(t) &= [\mathbf{w}^T(t), \mathbf{w}^T(t-1), \ldots, \mathbf{w}^T(t-N)]^T \end{aligned} \tag{4}$$

We can re-write the convolutive relationship (1) in the following static form:

$$\mathbf{y}_N(t) = \mathbb{J}_N\mathbf{s}_N(t) + \mathbf{w}_N(t) \tag{5}$$

using the signal $\mathbf{s}$ and the channel impulse response $\mathbf{h}$

$$\begin{aligned} \mathbf{s}_N(t) &= [\mathbf{s}(t), \mathbf{s}(t+1), \ldots, \mathbf{s}(t-N-L)]^T \\ \mathbf{h}_L &= [\mathbf{h}^T(0), \mathbf{h}^T(1), \ldots, \mathbf{h}^T(L)]^T \end{aligned} \tag{6}$$

and $\mathbb{J}_N(\mathbf{h}_t)$ is the generalized Sylvester resultant matrix [1];

$$\mathbb{J}_N = \begin{bmatrix} \mathbf{h}(0) & \cdots & \mathbf{h}(L) & 0 & \cdots & 0 \\ 0 & \mathbf{h}(0) & \cdots & \mathbf{h}(L) & \cdots & 0 \\ \vdots & & \ddots & \cdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{h}(0) & \cdots & \mathbf{h}(L) \end{bmatrix} \tag{7}$$

which is the block-Toeplitz matrix with $(N+1)$ rows of $(N+L+1)$ M-variate vector blocks, so that the dimension of this matrix is $(N+1)M \times (N+L+1)$.

Assuming $\mathbf{s}(t)$ to be a stationary signal with

$$\mathcal{E}\{\mathbf{s}_N(t)\mathbf{s}_N^H(t)\} = R_S \tag{8}$$

and taking into account the white noise properties, we get

$$R_y = \mathcal{E}\{\mathbf{y}_n(t)\mathbf{y}_n^H(t)\} = \mathbb{J}(\mathbf{h}_L)R_S\mathbb{J}_N^H(\mathbf{h}_L) + \sigma^2 I_{M(N+1)} \tag{9}$$

The problem of SIMO channel "blind" identification is that given a certain observation interval $\{t, \ldots, t+\tau_O\}$ at the output of the M-element antenna array, one has to estimate the Sylvester matrix (*i.e.* to estimate the $(L+1)\times M$-variate vector $\hat{\mathbf{h}}_L$ in (6)) or to estimate both $\hat{\mathbf{h}}_L$ and the actual transmitted waveform $\hat{\mathbf{s}}_{N_O}$.

For geolocation purposes, each of the (non-null) M-variate vectors $\hat{\mathbf{h}}(j)$, $j = 0, \ldots, L$ should then be decomposed in a number of (un-resolved in time) dominant plane-wave modes:

$$\mathbf{h}(j) = \sum_{k=1}^{K_j} a_{jk}\mathbf{s}(\hat{\theta}_{jk}) + \eta_j \tag{10}$$

where η_j is a permissible diversion from the plane-wave module, caused by "micro-multipath" ionospheric propagation via very closely spaced paths and/or cause by additive noise inflicted errors in vector $\hat{\mathbf{h}}(j)$ estimation.

One can see that in the general case, the problem solution requires order L estimation, followed by vector $\hat{\mathbf{h}}(j)$ and $\hat{\mathbf{s}}_{N_O}$ estimation, and then the compressive sensing type DoA estimation given in (10). At any stage in this estimation process, one may get an intermediate result which needs to have its validity

978-1-4673-3146-3/13/$31.00 ©2013IEEE

verified. Short of determining (in general computationally unachievable) global optimality in some statistical sense, one needs a means of doing a "quality assessment" of the order and vector estimates. To provide this, we wish to extend the Expected Likelihood (EL) test used in DoA estimation [2] to the problem at hand, as addressed in the next section.

II. Expected Likelihood Methodology for FIR SIMO Identification

Based on the above introduced model, the accurate statistical description, based on conditional (deterministic) maximum likelihood, available for the $M(N_O+1)$-variate observation vector is:

$$\omega_C[\mathbf{y}_{N_O}(t)] = \frac{1}{\pi^{M(N_O+1)}\sigma^{2M(N_O+1)}} \exp\{-\frac{1}{\sigma^2}\times \times (\mathbf{y}_n - \mathbb{J}_{N_O}(\mathbf{h}_l)S_{N_O})^H(\mathbf{y}_n - \mathbb{J}_{N_O}(\mathbf{h}_l)S_{N_O})\} \quad (11)$$

For the case with $MN_O > L + N_O + 1$ and the white noise power σ^2 known *a priori*, one can transform the CML case likelihood function (LF) given in (11) into the following likelihood ratio (LR):

$$LF_C(\hat{\mathbf{h}}_L) = \frac{[(M-1)(N_O-1) - \hat{L}]\sigma^2}{\text{Tr }\{P_\perp(\hat{\mathbf{h}}_L)\mathbf{y}_{N_O}(t)\ \mathbf{y}^H_{N_O}(t)\}} \quad (12)$$

where

$$P_\perp(\hat{h}_L) = I_{M(N_O+1)} - \mathbb{J}(\hat{\mathbf{h}}_L)[\mathbb{J}^H(\hat{\mathbf{h}}_L)\mathbb{J}(\hat{\mathbf{h}}_L)]^{-1}\mathbb{J}^H(\hat{\mathbf{h}}_L) \quad (13)$$

while the signal can be estimated as

$$\hat{S}_{N_O} = [\mathbb{J}^H(\hat{\mathbf{h}}_L)\mathbb{J}(\hat{\mathbf{h}}_L)]^{-1}\mathbb{J}^H(\hat{\mathbf{h}}_L)\mathbf{y}_{N_O}(t) \quad (14)$$

It can be demonstrated that for the clairvoyant solution $\hat{\mathbf{h}}_L = \mathbf{h}_L$, the probability density function (p.d.f.) for $LF_C(\mathbf{h}_L)$ does not depend on $\mathbf{h}_L$, though the order L of the model does need to be known [2]. This forms the basis of the Expected Likelihood technique, which uses quality assessment metrics from this likelihood ratio distribution, as it is independent of any particular scenario realization.

From (11)-(14), it follows that the entire set of (N_O+1) M-variate observation vectors $\mathbf{y}(t),\ldots,\mathbf{y}(t+N_O)$ should be treated as a single (N_O+1)-variate "snapshot" to be properly statistically described. In general, any partition of this single snapshot into a set of $M(N+1), N < N_O$-variate training samples (the technique used in all existing SIMO blind identification techniques [3, 4]) leaves these vectors correlated, violating the i.i.d. assumption inherent in (12).

Of course, if we assume the source signal $\mathbf{s}(t)$ to be temporally white, *i.e.* $R_S = \sigma_S^2 I_{N+L+1}$, and allow for longer than $m > L$ "time gaps" between the successive $M(N+1)$-variate training samples, we may enforce the independence of these training samples, but the total number of such i.i.d training samples is then only equal to $N_O/(N+L+1)$, which is significantly less than the number of training samples, that one obtains using the conventional "sliding window" averaging $(T = N_O - N)$ [5]:

$$\hat{R}_y = \frac{1}{T}\sum_{t=N+1}^{N+T} \mathbf{y}_N(t)\mathbf{y}_N^H(t) \quad (15)$$

Specifically, this sample covariance matrix is used by many existing methodologies (subspace-based [3], linear prediction [4], and others) as the input statistics for identification. Therefore validity of any estimated covariance matrix model

$$\hat{R}_y = \mathbb{J}_N(\hat{\mathbf{h}}_L)\hat{R}_S\mathbb{J}_N^H(\hat{\mathbf{h}}_L) + \sigma^2 I_{M(N+1)} \quad (16)$$

should be tested with respect to this covariance matrix. Of course, when identification is completed, with $\mathbf{h}_L$ (or in fact $(\hat{\theta})_{jk}, j = 0,\ldots,L$ in (10) being identified, one can apply the CML test in (12). Yet a number of important decisions have to be performed prior to the final model being available.

Note that temporal correlation between "siding window" training samples not only invalidates the formation of the traditional likelihood function based on i.i.d. assumptions of the training samples, but also means that the sample matrix $\hat{R}_y$ in (15) is not described by the complex Wishart distribution $\mathcal{CW}(T, M(N+1), R_y)$.

Regardless, the mean value of the sample matrix $\hat{R}_y$ in (15) is equal to the actual covariance matrix R:

$$\mathcal{E}\{\hat{R}_y\} = \frac{1}{T}\sum_{t=N+1}^{N+T}\mathcal{E}\{\mathbf{y}_N(t)\mathbf{y}_N^H(t)\} = R_y \quad (17)$$

and therefore the covariance matrix model $R_y(\hat{\mathbf{h}}_L)$ may be treated as appropriate if

$$\mathcal{E}\{R_y^{-\frac{1}{2}}(\hat{\mathbf{h}}_L)\hat{R}_y R_y^{-\frac{1}{2}}(\hat{\mathbf{h}}_L)\} = I_{M(N+1)} \quad (18)$$

For this reason, we may still consider as a "quality metric" likelihood ratios that have been derived under the i.i.d. assumption, even if not strictly accurate. In particular, we may use the sphericity test to check the hypothesis [6]:

$$\begin{aligned} H_0 &: \ \mathcal{E}\{\hat{R}_y\} = cR_y(\hat{\mathbf{y}}_L) \ \text{ versus} \\ H_1 &: \ \mathcal{E}\{\hat{R}_y\} \neq cR_y(\hat{\mathbf{y}}_L) \ \ c > 0 \end{aligned} \quad (19)$$

when

$$LR_{sp}(R_y(\hat{\mathbf{h}}_L)|\hat{R}_y) = \frac{\det[R_y^{-1}(\hat{\mathbf{h}}_L)\hat{R}_y]}{\left[\frac{1}{M(N-1)}\text{Tr }[R_y^{-1}(\hat{\mathbf{h}}_L)\hat{R}_y]\right]^{\frac{1}{M(N-1)}}} \quad (20)$$

and look for the model $R_y(\hat{\mathbf{h}}_L)$ that maximizes this LR_{sp} value. For "Expected Likelihood" (EL) applications, this criterion is suitable only if one can demonstrate for the H_0 hypothesis when $R_y(\hat{\mathbf{h}}_L) = R_y(\mathbf{h}_L)$ that the p.d.f. of $LR_{sp}[R_y(\mathbf{h}_L|\hat{R}_y]$ does not depend on the unknown parameters $\mathbf{h}_L$, and is therefore fully specified by *a priori* known parameters such as M, N, and T. In this case, the p.d.f. can be precalculated and metrics derived from that p.d.f. can be used as "quality assessment" thresholds for the EL methodology. There, a model $R_y(\hat{\mathbf{h}}_L$ is treated as appropriate if the likelihood ratio $LR_y[R_y(\hat{\mathbf{h}}_L)|\hat{R}_y]$ generated by this estimated covariance matrix model is above (or within) the specified "quality assessment" thresholds.

Note that individually, each vector

$$\xi(t) = \{R_y^{-\frac{1}{2}}(\mathbf{h}_L)\mathbf{y}_N(t)\} \sim \mathcal{CN}(0, I_{M(N+1)}) \quad (21)$$

and therefore $\eta \equiv LR_{sp}[R_y(\mathbf{h}_L)|\hat{R}_y]$ may be presented as

$$\eta = \frac{\det\{\frac{1}{T}\sum_{t=N+1}^{N+T}\xi(t)\xi^H(t)\}}{\left\{\frac{1}{M(N-1)T}\sum_{t=N+1}^{N+T}\xi(t)\xi^H(t)\right\}^{\mathbf{y}_{M(N-1)}}} \qquad (22)$$

We end this section with our core conjecture for application of Expected Likelihood to the FIR SIMO identification problem:

> Conjecture: A metric η using the p.d.f. of $LR_{sp}[R_y(\mathbf{h}_L)|\hat{R}_y]$ calculated for the noise-only case (*i.e.* for $h_L = 0$ or $R_S = 0$) may be used as a p.d.f. for η for an arbitrary $R_y(\mathbf{h}_L) \neq I_{M(N+1)}$ scenario with sufficient for practical application accuracy.

Since we have not proven this conjecture, in the next section we present simulation results which establish this conjecture for several representative scenarios.

III. Simulation Results

We consider a single-site location (SSL) direction-finding case, where a $M = 16$-element uniform line array (ULA) observes an HF communications signal that propagates over three resolved in time ionospheric propagation modes $\tau = 0, 1, 3$ and therefore span $L = 3$ intervals with $L+1 = 4$ overall number of "taps" in a corresponding finite impulse response (FIR) filter. The internal antenna noise is white with unit power ($\sigma^2 = 1$), while the source is simulated as an autoregressive (AR(1)) process with

$$R_S = \sigma_S^2|\rho^{|p-q|}|,\ p,q \in 0,\ldots,N+T+1. \qquad (23)$$

The ratio σ_S^2/σ^2 denotes the SNR (per element) in our simulations. Following our methodology in (5)-(7), the Sylvester $M(N+1), N+1$ matrix is formed using three non-zero vectors $\mathbf{h}(0), \mathbf{h}(1), \mathbf{h}(3)$:

$$h(j) = \sum_{i=1}^{2}\alpha_{ij}\mathbf{s}(\theta_{ij},\ \alpha_{ij}) \sim \mathcal{CN}(0,1) \qquad (24)$$

and specific DoA values as introduced in Table I. Moreover, each vector $\mathbf{h}(j)$ is scaled by a corresponding propagation factor that account for different gains of the considered propagation modes.

As follows from Table I, modes with different delays arrive at quite different elevation angles, while specific delays in turn consist of two poorly resolved in elevation plane waves. Yet the main purpose of our simulations is to demonstrate validity of the introduced Expected Likelihood principle, rather than efficiency of any particular identification algorithm. Therefore the sensitivity of our study to these angular parameters should be essentially zero, and accuracy of their estimation is not pursued further in this study.

For the fixed $\mathbb{J}_N(\mathbf{h}_L)$, we conducted 10^6 independent Monte-Carlo trials for $T = 1000$ and calculated sample p.d.f.s for $LR_{sp}[R_y(\mathbf{h}_L)|\hat{R}_y]$ for different parameters of the problem. First, in Fig. 1, we see demonstrated the LR distribution invariance with scenario parameters suggested by our Conjecture. In Fig. 2, the number of propagation modes were varied and again, the LR distributions demonstrate scenario invariance.

TABLE I. Simulation Parameters

Item	Symbol	Value
Array Size	M	16
Time Samples	T	1000
Tap Gain 1-4	$\sigma_S(1 = 4$	10, 5, 0, 2;
Signal Power	σ_S^2	1, 0.15, or 0.05
Noise Power	σ^2	1
Temporal Corr	ρ	0, 0.5, 0.9 or 0.99
El Angle Pair 1 1	$\theta(1,2)$	10 and 12 †
El Angle El Angle 2	$\theta(3,4)$	30 and 31 †
El Angle El Angle 3	$\theta(5,6)$	60 and 61 †
El Angle El Angle 4	$\theta(7,8)$	70 and 70.5 †

†The 4 main sources resolve in time and angle - Unresolved sources have only a small angle offset which does not resolve in elevation angle. This offset varies with elevation angle because the elevation resolving power of the array improves toward zenith.

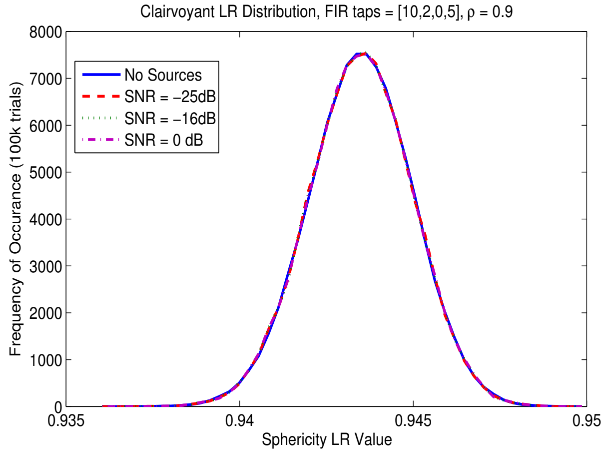

Fig. 1. Sphericity test LR distribution for noise only and three source scenarios (SNRs of -25dB, -16dB, and 0dB). Temporal correlation ρ was set to 0.9. The LR p.d.fs show a "practical" scenario invariance property which is a key requirement for the application of Expected Likelihood.

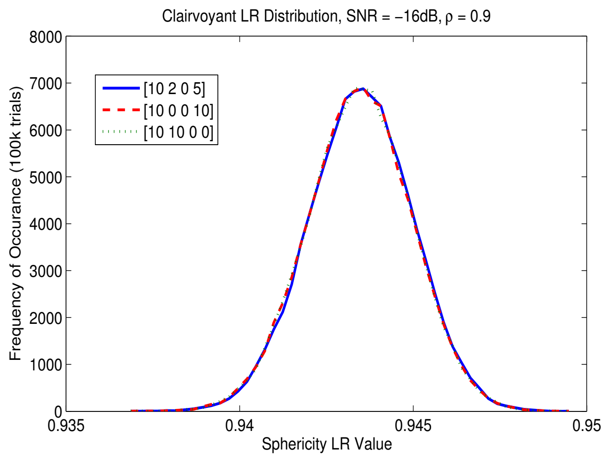

Fig. 2. Sphericity test LR distribution for three source scenarios, all with SNR of -16dB, but with propagation coefficients of [10; 2; 0; 5], [10; 0; 0; 10], and [10; 10; 0; 0]. This effectively gives scenarios with 3 modes, two widely spaced (in delay) modes, and two closely spaced (in delay modes). The LR p.d.fs continue to show a practical scenario invariance property.

In Fig. 3, we introduce the sample p.d.fs. calculated for the base model (SNR = -16dB, tap gains = [10; 2, 0. 5]) and introduce different temporal correlations for our source signal. One can see that temporal correlation of the source signal also does not affect the properties of the "Expected Likelihood" distribution $\omega(\eta)$ in (22) until introduction of very extreme temporal corrlations ($\rho > 0.99$) and even then, there is only a minor offset which would still allow use of the EL methodology, albeit with some minor false alarm elevation.

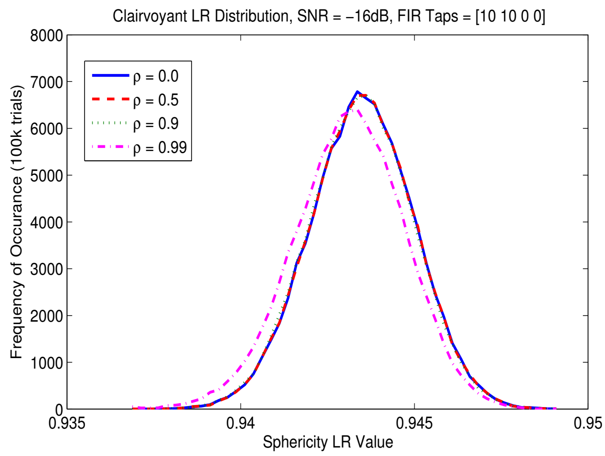

Fig. 3. Sphericity test LR distribution for the same source scenario -16dB and propagation coefficients of [10; 2; 0; 5]. The LR p.d.fs continue to show a practical scenario invariance property, except under the most extreme temporal correlations.

It is also important to note that the demonstrated invariance of the $\omega(\eta)$ does not mean that the vectors $\xi(t)$ in (22) may be treated as independent. At Fig. 4, we once again introduce sample p.d.f.s for "sliding window" produced noise samples as in (22) and the traditional sphericity test p.d.f. for the same number $T = 1000$ of i.i.d. noise-only training samples. It is remarkable that the mean value for both p.d.f.s remains the same, while the greater number of i.i.d. samples in the traditional case leads to a considerably smaller second order central moment.

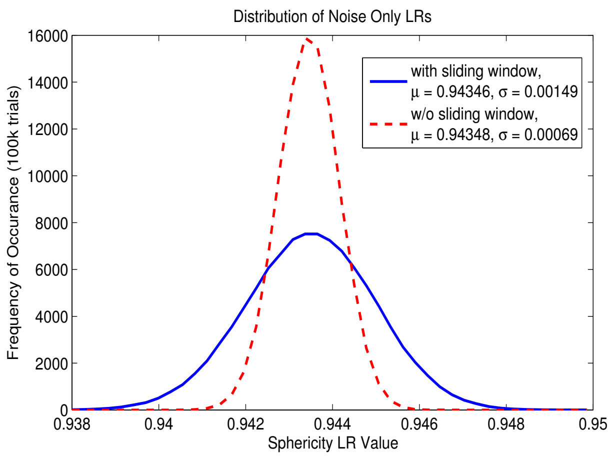

Fig. 4. Sphericity test LR distribution for a noise only source scenario utilizing a sliding window and no sliding window for the sample data generation. The means of the distributions correspond, but the second moments do not.

It is also important that if one tries to enforce the independence of training samples by appropriate "sparsing" of training data, the efficiency of the EL quality assessment will be significantly degraded. Simulation employing a noise-only scenario with the traditional i.i.d. sample environment with 145 training samples rather than the sliding window with 1000 training samples was used to generate sample LR p.d.f.s. Obviously, for $M(N+1) = 112$, this relatively small sample support leads to significantly smaller likelihood ratio values (specifically in this case, values ranging between 0.545 to 0.57). Practically, it means that sliding window produced training data collectively contain much more information about the underlying scenario than the properly sub-sampled strictly i.i.d. subset, and therefore these sliding samples should not be discarded neither for identification nor for quality assessment purposes.

IV. Conclusion

We have suggested for the FIR SIMO identification problem that likelihood ratio tests (a sphericity test in particular) derived for i.i.d. training samples (even though that condition that does not apply for the FIR SIMO scenario) can still be used in lieu of a fully accurate likelihood function accounting for the correlation of the training samples. It has been conjectured and demonstrated by simulation that the p.d.f. of the sphericity test for the true space-time covariance matrix can be accurately represented by the p.d.f. of the LR for spatially white noise-only samples, derived by a "sliding window" technique from a single white noise sequence. While accurate analytical expressions for this p.d.f are not derived, independence of this "Expected Likelihood" p.d.f on parameters other than *a priori* known ones enable use of the EL methodology.

Specifically, an estimated space-time covariance matrix model is treated as acceptable, if sphericity test values generated by this model is within the support of this EL distribution. In this way, the derived soution is statistically at the same likelihood "distance" from the sample matrix as the unknown true solution, which means that further model refinement is not statistically necessary, and the model is "as likely as the truth".

Future activity can be focused on proving the conjectured statistical properties and analysis of EL efficiency for identification solution quality assessment and determination of estimate breakdown in threshold conditions.

References

[1] R. Bitmead, S. Kung, B. Anderson, and T. Kailath, "Greatest common divisor via generalized Sylvester and Bezout matrices," *IEEE Trans. Automat. Control*, vol. 23, no. 6, pp. 1043–1047, 1978.

[2] B. A. Johnson and Y. I. Abramovich, "DOA estimator performance assessment in the pre-asymptotic domain using the likelihood principle," *Elsevier Journal of Signal Processing*, vol. 90, no. 5, pp. 1392–1401, May 2010.

[3] K. Meraim, J. Cardoso, A. Gorokhov, P. Loubaton, and E. Moulines, "On subspace methods for blind identification of single input-multiple output FIR systems," *IEEE Trans. Signal Processing*, vol. 45, no. 1, pp. 42–55, Jan 1997.

[4] A. Gorokhov and P. Loubaton, "Blind identication of MIMO-FIR systems: A generalized linear prediction approach," *Elsevier Journal of Signal Processing*, vol. 73, pp. 105–124, 1999.

[5] C. B. Papadias and D. Slock, "Normalized sliding window constant modulus and decision-directed algorithms: A link between blind equalization and classical adaptive filtering," *IEEE Trans. Signal Processing*, vol. 45, no. 1, pp. 231–235, 1997.

[6] R. Muirhead, *Aspects of Multivariate Statistical Theory*. New York: Wiley, 1982.